REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-04-
0163

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructi
data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspec
this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188),
4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to c
valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
09Feb2004

2. REPORT TYPE
Final Technical

DATES COVERED *(From - To)*
15SEP2000 to 14SEP2003

4. TITLE AND SUBTITLE

MURI Fellowship: Fundamental Principles in Adaptive Learning Technology

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-00-1-0381

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Shebilske, Wayne L.
Gildea, Kevin M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research and Sponsored Programs
Wright State University
3640 Colonel Glenn Hwy.
Dayton, OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL
Office of Scientific Research
4015 Wilson Blvd. Rm 713
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approve For Public Release: Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This grant was a fellowship to support a graduate student, Kevin Gildea, who participated in a Multi-University Research Institute Grant, which focused on the development of distributed intelligent agents to support the creation of new individual and team training protocols, along with the rigorous experimental testing of the new methodologies.

20040319 118

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					19b. TELEPHONE NUMBER *(include area code)*

AFOSR Final Report

Wayne L. Shebilske

Kevin M. Gildea

1 Introduction –

This grant was a fellowship to support a graduate student, Kevin Gildea, who participated in a Multi-University Research Institute Grant, which focused on the development of distributed intelligent agents to support the creation of new individual and team training protocols, along with the rigorous experimental testing of the new methodologies.

Support from the AFOSR fellowship and from the MURI grant enabled Kevin Gildea to participate in many experiments related to the MURI grant and to take a lead role in completing two experiments that were both important to the MURI projects and foundations for his Master's thesis and his doctoral dissertation. His Master's thesis paved the way for developing and testing a revised Space Fortress (RSF) research tool. It also integrated previously separate research tracks. One track was on the relationship between performance and attribution of stressors as challenges or threats; the other track was on training complex skills. Predicting performance is critical in both tracks. In his Master's thesis, Kevin Gildea showed that trainees perceive the stress of leaning a new complex skill as either a challenge or a threat. Those who perceive a challenge perform better than those who perceive a threat. The advantage for those who felt challenged was present during acquisition, retention, and transfer of training. In his doctoral dissertation, Kevin Gildea extended this finding to team training for a Dynamic Distributed Decision (DDD) making task. He made two other important contributions. The second contribution was redesigning the DDD task so that teams who did better teamwork accomplished their mission better. The third contribution was developing and successfully testing an alternative to cross training for teams. The alternative is an Active Interlocked Modeling (AIM)-Dyad training protocol, in which each teammate has a training partner with whom the task is divided during practice. During tests, all teams perform without a partner under identical conditions. Kevin Gildea hypothesized that the AIM-Dyad protocol would reduce workload during practice and increase learning of team knowledge and team skills and that improved team skills would improve team performance. Preliminary results support the hypothesis. Gildea completed data collection for his dissertation in February before accepting a position with Aptima, the developers of the DDD task, to continue his work on refining procedures for experiments with DDD. Completion of his dissertation and his doctoral graduation is anticipated this spring.

Gildea's research activities will be reviewed briefly in the following sections.

2 Research Activities

2.1 Testbed Development – Revised Space Fortress (RSF)

Kevin Gildea participated in extensive validation tests of RSF, which was subsequently used in other experiments. One goal in developing RSF was to provide more data without changing the task. Another goal was to enable extensions. One extension was the addition of *hierarchical decomposition tests*, (suggested initially by Fredrickson and White (Frederiksen and White 1989)). A second extension was the addition of Adaptive Multiple Emphasis on Components (AMEC) variations. This spring, our team expects to release the software for RSF to others who wish to use it; we have already been in contact with several researchers who would like to use it.

2.1.1 Validation Testing and Results

The validation tests were conducted on three levels. The focus of the lowest level unit testing was on confirming that the large number of parameters underlying the simulation did indeed match those of SF. These included such things as object dimensions (ship, fortress, hexagons, mines, shells, and missiles), maximum object speeds (ship, missiles, shells, mines), frequencies at which entities like mines and bonuses appear, etc.. While the initial values were determined by reading the SF code, measurements were taken on the screen during a trial to identify any incorrect values that might have been accidentally entered. A variety of measurement techniques were used to obtain suitable measurements. A complete discussion of the parameters checked, the measurement techniques used, and the corrections made are in Johnson et al., submitted.

Next, we had expert SF performers execute both simulations and report any discrepancies they noted. Kevin Gildea was one of these experts. The purpose in these tests was to discover any methodological differences that might have crept into the coding. One minor difference was observed. Experts felt that the joystick control in RSF was very slightly more sensitive than that of SF. We believe this is due to the underlying joystick reading routines, which were necessarily different in the two systems. This was one of two things that led to a hypothesis about expected differences in performance (discussed briefly below).

Then, we performed extensive timing tests to determine if ultimately, we were able to obtain adequate real-time performance. It was also of interest to observe the growth in memory utilization by the SimEngine during a trial, since this is related the garbage collection times. Both tests yielded good results (see (Johnson et al., submitted)).

Finally, we conducted human subject experiments to compare the results of the original PC version of Space Fortress with RSF. However, during the validation process, we discovered a scoring error in the original PC version of SF. In particular, scores for flying outside of large hexagon were not penalized as they should have been. This, in conjunction with the slight sensitivity difference led to some hypotheses about differences in the results of using RSF and SF. Most notably, we predicted that, relative to participants training with SF, participants training with RSF would have more of a tendency to fly within the hexagon after training (because the scoring rewarded doing so). As a result, they would score worse in the beginning of training, especially on control score, and better in most or all scores by the end of training, though there is no reason to believe they would perform better on the control score. On crossover tests, participants who trained with RSF would maintain their advantage when they switched to SF, because they would have the habit of flying within the hexagon and would not experience the consequences of the scoring error in SF. In contrast, participants who trained with SF would fall even lower when they switched to RSF, because they would have the habit of flying outside the hexagon, which would reduce their control score in the correct RSF scoring system. These hypotheses were largely substantiated in our experiments. For example, Figure 1 shows a comparison of Total Scores, and Figure 2 shows a comparison of the Control Scores. Kevin Gildea was a co-author on the article in which the details are described (Shebilske, Volz et al. submitted).

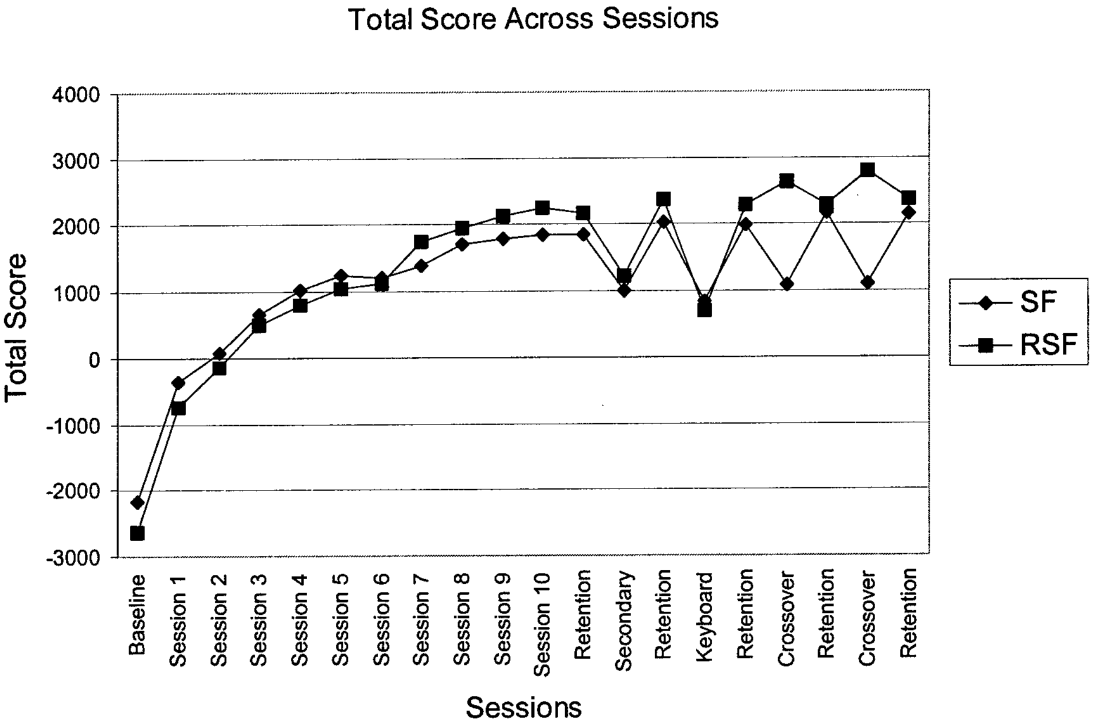

Figure 1. Total Score comparison.

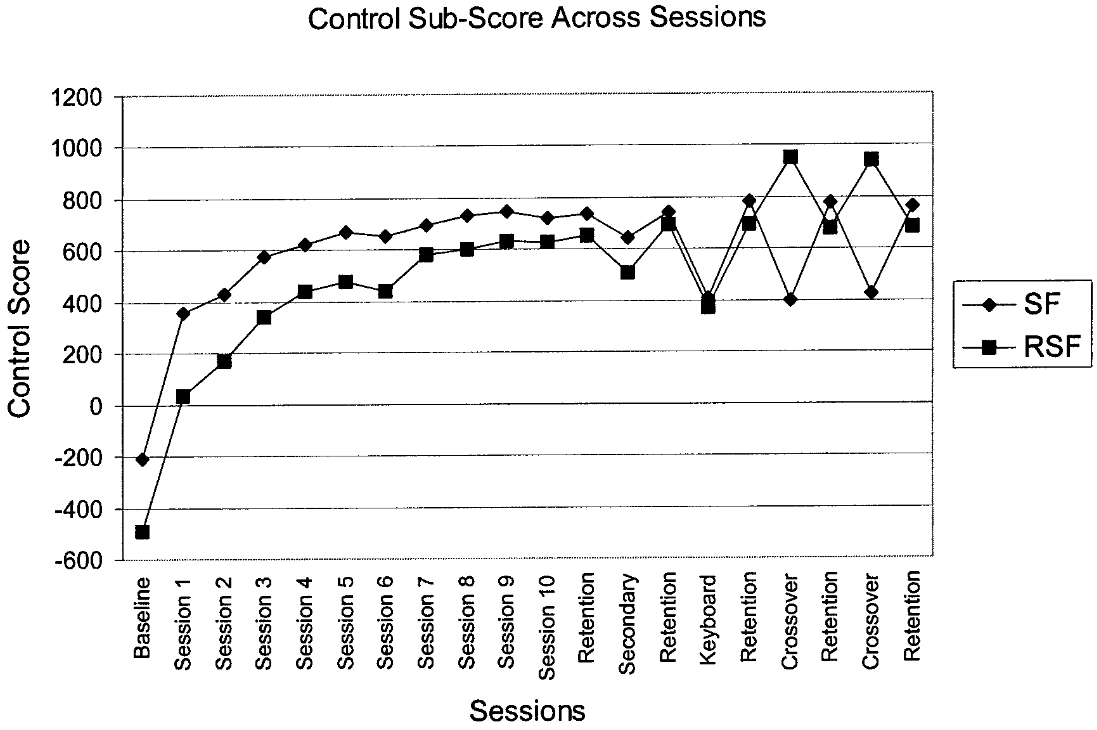

Figure 2. Control Score comparison

As there are minor differences between the results of SF and RSF, one must be cautious in comparing the results of using the two. However, the disadvantage of having to be cautious when comparing results between SF and RSF is offset by many advantages of RSF. The advantages include reducing errors in SF, running on current generation platforms, support for team training, support for incorporation of intelligent agents, flexible experiment definition, global management of experiment data, and storing enough data to playback each trial for detailed analysis. The accommodation of intelligent agents enables dynamic adaptive interactions, which allows, for example, simulating environmental dynamics, coaches, and teammates, and/or adapting training protocols to individual or team performance. Although the playback feature is currently available, it was added after this experiment, so we could not use it to supplement our analysis of predicted differences. It would have been interesting to play back trials to directly investigate predicted differences in the development of the strategy of circumnavigating the Fortress. Such investigations will be possible in future experiments.

2.1.2 Extrinsic Hierarchical Decomposition (HD) Tests

Frederiksen and White (1989) developed training modules for Space Fortress based on a hierarchical task decomposition. The present HD tests were based on these training modules. One important difference is that each task is repeated few times for tests, as opposed to many times for training. Another important difference is that performance on each test is scored more precisely. For example, errors in making a 45-degree turn are measured in degrees of deviation from 45 degrees. The following extrinsic HD skills tests have been developed and incorporated into RSF:

1. Firing on the Fortress
2. Aiming the Spaceship while it is Stationary
3. Slowing Down and Stopping the Spaceship
4. Aiming a Moving Spaceship from Different Distances
5. Aiming and Firing from Different Distances
6. Aiming and Firing from Different Trajectories
7. Making a 45 Degree Change in Trajectory
8. Slowing Down While Turning
9. Turning Through Different Angels by Varying the Thrust
10. Controlling the Speed and Trajectory of the Spaceship
11. Changing Trajectories at Corners of a Hexagon
12. Navigating the Hexagon while Aiming at the Fort
13. Navigating the Hexagon While Firing
14. Navigating about the Fortress while it is Tracking
15. The Full Game Without Mines and Resources

Kevin Gildea participated in preliminary experiments with these HD tests.

2.1.3 AMEC Test Suite

Multiple Emphasis on Components (MEC) is a training protocol in which trainees perform the whole task at all times while focusing attention on task components (e.g. Gopher, Weil, & Bareket, 1994). The protocol facilitates average and poor trainees, but not good trainees. A possible reason is that MEC advances all trainees from one component to the next at the same time. We hypothesized that this advancement schedule may be appropriate for average and poor trainees, but is too slow for good trainees. We developed an adaptive MEC (AMEC) to test this hypothesis. Our original AMEC was moderately successful, but preliminary experiments with it suggested the need for changes in a new AMEC.

The test suite for new AMEC protocols includes three features that are new with respect to our original AMEC:

1. Forced advancement
2. Explicit numeric goals
3. Learning focus

The original AMEC had none of these features. The new AMEC protocols have forced advancement. This feature advances trainees to the focus that they would have had in MEC regardless of whether they had reached their goal for the previous focus. Forced advancement prevents low ability trainees from being denied the opportunity to practice basic skills in the context of focus on more advanced skills. Forced advancement has the effect of defaulting to an MEC protocol, which is known to work well for low ability trainees. The forced advancement protocol remains adaptive in the sense that a trainee can move ahead of the MEC schedule whenever they achieve the goal for the current and all previous foci.

Explicit numeric goals inform the trainee of numeric goals for each sub-score on which they focus. A pilot experiment indicated that trainees did not attend to the goals when they had to compare their memory of the goal with sub-scores in the SF display. Therefore the explicit numeric goals were implemented by means of dynamic bar graphs near the SF displays. A dynamic bar changed height moment to moment to indicate changes in the focal sub-score. A line above the bar indicated the goal for the sub-score. If trainees achieved the goal on the current focal sub-score but fell below their goal on a previous focal score, bar graphs would appear for all previous sub-scores. In addition, trainees were told that they had reached their goal on the current focal sub-score, but they had backslid on their previous goals. They were also told that they had to score at or above the goal on the current and past sub-scores before they could advance.

The learning focus was implemented through instructions that were identical to those in AMEC except for the additional statement that the trainee should focus on a sub-score in order to learn the skills necessary for that sub-score.

When one considers these three features in combination, there are eight different possible combinations. These have all been implemented in RSF and Kevin Gildea has participated in preliminary experiments with them.

2.2 Appraisal of Stress Related to Training Complex Skills

In his Master's thesis, Gildea (2001) noted a relationship between a defining characteristic of complex skills and a predictor of performance in the literature on stress. The defining characteristic is that the tasks overwhelm a trainee (cf. Schneider, 1985). Kevin Gildea reasoned that if one views being overwhelmed as stressful, then one can predict training performance based on whether trainees view the stress as a challenge or a threat. He tested this hypothesis using a standard 8-item stress appraisal scale during the development and testing of RSF. The results supported his hypothesis. By the end of training, trainees who felt challenged performed almost twice as well trainees who felt threatened, and they maintained this advantage on test of retention and transfer. This result has implications for screening trainees or for developing protocols that foster challenge instead of threat.

In his doctoral dissertation, Kevin Gildea found that trainees who felt challenged by a DDD AWACS simulation performed significantly better than trainees who felt threatened.

2.3 Teamwork and Team Performance in a DDD AWACS Simulation

Our research team's first DDD experiment replicated and extended procedures and scenarios developed by Hollenbeck et al. (2002) to study the relationship between personality and helping on teams. They provided the software, and details about the procedures. Our team replicated the procedures as closely as possible including the exact scenarios and instructions in one condition, which instructed teammates to help one another as much as possible (high help). We extended the procedure in another condition, which instructed teammates to help one another as little as possible (low help). Half of the teams had the high help instruction first and half had the low help instruction first.

Hollenbeck et al. (2002) reported that teammates who were high in extroversion were more likely to request help and to provide help. In addition, teammates high in conscientiousness modulated their requests for help with respect to whether or not they had enough resources to cope with a high load without help. Hollenbeck et al. (2002) reported neither individual nor team performance.

We selected their paradigm over several other AWACS simulations because the procedures were adapted to college students while others were better suited for trainees with more military experience. For example, in an unpublished validation study of another AWACS simulation, Barry Goettl and his colleagues compared actual AWACS instructors with actual trainees for AWACS. The experimenters provided five eight-hour days of pre-training to trainees before they compared their performance with instructors (personal communication). In contrast Hollenbeck et al. (2002) collected meaningful data from college students, each of whom participated for one eight-hour day.

We measured the big five personality factors to investigate the relationship between personality and helping behavior, which we operationally defined as the number of assist attacks (attacking an enemy track in a teammates quadrant), and number of Id transfers (electronically communicating track Ids to teammates). We also measured team offensive score and team defensive score, which are standard DDD measures of mission success, in order to investigate the relationship between helping and team performance.

Investigating this relationship was important to us because our interviews with subject matter experts suggested that high helping is generally associate with high team performance in operational settings. Accordingly, we planned to develop Intelligent Agents that would facilitate training of helping behaviors and to test those agents in a laboratory task.

Our results suggested that the original Hollenbeck et al. (2002) procedures were useful for their intended purpose of studying the relationship between personality and helping, but the procedures were not appropriate for studying the relationship between helping and team performance.

This report will focus on our manipulation of low and high helping instructions to investigate the relationship between helping and team performance. Two manipulation checks indicated that the instructions affected helping as intended. First, the number of assist attacks per mission was 16 in the low help condition and 24 in the high help condition, $\underline{F}(1,38) = 50.36$, $\underline{p} < .0001$. Second, the number of Ids transferred per mission was 132 in the low help condition and 142 in the high help condition, $\underline{F}\ (1,38) = 5.95$, $\underline{p} < .05$. However, increased helping did not increase team performance. In fact, the team offensive score went in the opposite direction. The team offensive score was 1069 in the low help condition and 1046 in the high help condition, $\underline{F}\ (1,38) = 8.48$, $\underline{p} < .01$. The team defensive score did not change significantly. It was 34718 for the low help condition and 34751 for the high help condition, $\underline{F}\ (1,38) < 1$.

2.4 Redesigning DDD Procedures and Scenarios

These results motivated us to revise the procedures to create a laboratory analogue of the conditions in operational settings in which high helping is associated with high team performance. Our revisions were guided by two goals. The first was to increase the level of teamwork skills and the second was to change the scenarios to afford advantages to teams with better teamwork. The goal of improving team skills emerged from recordings, which indicated poor team skills (e.g. little communication and coordination). We attributed the poor team skills to an important difference between the original Hollenbeck et al. (2002) procedures with those employed in operational settings and in other AWACS simulations. Goettl and his colleagues not only provided five days of pre-training, they also provided a planning session and before each mission and a debriefing session after each mission. Checklists similar to those used by actual AWACS teams were provided to guide the planning and debriefing sessions. Our revision, therefore, added planning and debriefing sessions with similar checklists. Our second goal was to develop scenarios that afforded an advantage to teams with better teamwork. We discovered that in the original scenarios, teams could do as well or better by exclusively defending their own quadrant as they could by helping one another.

Kevin Gildea played an important role in revising the procedures. He developed scenarios in which the enemy tracks stayed within the quadrant they entered, and were equal in number and power to AWAC team's vehicles. This equivalence meant that an AWACS team could defend against all the enemy tracks. To do so, however, the teammates had to leave their quadrants to help other teammates. Kevin Gildea also helped develop an intelligent report and a planner that could be used during the planning session and the mission. During the planning session, trainees use the intelligence report and the planner to plan where they would position their vehicles to defend against the tracks in the intelligence report. During the mission, the trainees use the intelligence report and planner to stay as close as possible to the plan. Finally, Kevin Gildea helped articulate the checklists for planning and debriefing. One goal of the planning checklist was to convey the tradeoffs between preparing contingency plans, adapting to unexpected situations, and committing to a plan. Goals for the debrief included directing attention to actions that caused successes and failures and stimulating discussion of alternative plans that increase successes and decrease failures in future missions.

Analyses of results with the new procedures have not been completed. However, a preliminary median split of the available data suggests that teams that help more will score better on team offensive and defensive scores.

2.5 Aim-Dyad Protocol for Training in a DDD AWACS Simulation

Kevin Gildea also developed and tested an Aim-Dyad protocol for training in the revised DDD scenarios. In the standard control protocol, each teammate had four vehicles, an AWACS, a jet, a helicopter, and a tank during test and practice missions. In the Aim-Dyad protocol, each teammate had the same four vehicles an AWACS, a jet, a helicopter, and a tank during tests, but they had a partner who controlled half of the protocol would reduce workload during practice and increase learning of team knowledge and team skills, and that improved team skills would improve team performance. Preliminary results support the hypothesis.

2.6 Advancing Research on Intelligent Agents for Distributed Team Training

Kevin Gildea's research advanced the developing and testing of Intelligent Agents for distributed team training. Our research team refined partner agents and coaching agents in RSF before developing them in DDD. The validation of RSF was essential because it showed how results with RSF, which had an interface for intelligent agents related to results of SF, which did not. The tests of AMEC revealed the importance of specific goal related feedback in coaching agents. The Hierarchical Decomposition tests provided more fine-grained analyses of training with partner agents. The research on stress appraisal paved the way for developing agents that would reduce threat appraisals and increase challenge appraisals.

The analysis of teamwork, as reflected in helping, and team performance on missions indicated the need to revise our DDD procedures in order to accomplish our goal of providing a representative laboratory analogue for evaluating intelligent agents. The agents are designed to improve team performance by improving aspects of teamwork, such as helping. It is necessary, therefore, that there be a positive relationship between high helping and mission success. Our team waited to determine whether this necessary condition was present before implementing vital steps in developing intelligent agents, such as cognitive task analyses. These steps started immediately after the successful revision of DDD.

Finally, the successful development of an Aim-Dyad protocol for training teams in DDD is an important step in advancing partner agents from training individuals in RSF to training teams in DDD. A partner agent replaced a beneficial human partner in RSF. Similarly, developing a beneficial human partner for team training in DDD sets the stage for developing a partner agent to replace the human partner. This development will be guided by the analogous agent development in RSF. In DDD, the partner agents will be designed to reduce workload during practice in order to facilitate the learning of team skills.

3 Publications

Gildea, K. M. (2001). *Threat and challenge appraisal and learning a complex skill.* Unpublished master's thesis, Wright State University, Dayton, OH. (Manuscript in preparation for publication).

Johnson, J. C., Sen Cao, Maitreyi Nanjanath, Jonathan Whetzel, Thomas R. Ioerger, Barani Raman, Wayne Shebilske, and Dianxiang Xu, "Fine-Grained Data Acquisition and Agent Oriented Tools for Distributed Training Protocol Research: Revised Space Fortress," submitted to *Behavior Research Methods, Instrumentation & Computers*, June 6, 2003.

Shebilske, Wayne L., Richard A. Volz, Kevin M. Gildea, Judson Workman, Maitreyi Nanjanath, Sen Cao, and Jonathan Whetzel, "Revised Space Fortress: A Validation Study" submitted to *Behavior Research Methods, Instrumentation & Computers*, June 6, 2003.

4 References

Frederiksen, J. R. & White, B. Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 71,* 89-146.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36,* 387-405.

Hollenbeck, J. R., Moon, H., Ellis, A. P. J., West, B. J., Sheppard, L., et al. (2002). Structural contingency theory and individual differences: Examination of external and internal person-team fit. *Journal of Applied Psychology, 87,* 599-606.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27,* 285-300.